Hameed, N. et al. (2019). Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques. *Expert Systems with Applications*. 141, 112961.

Multi-Class Multi-Level Classification Algorithm for Skin Lesion Classification using Advanced Machine Learning Techniques

Nazia Hameed[a,*], Antesar Shabut[a], M. A. Hossain[a], Miltu K. Ghosh[a,b]

[a] Anglia Ruskin IT Research Institute (ARITI), Anglia Ruskin University, Chelmsford, UK

[b] Department of Pharmacy, NSHM Knowledge Campus, India

Emails:

Nazia Hameed: nazia.hameed@pgr.anglia.ac.uk
Antesar Shabut: antesar.shabut@anglia.ac.uk
M. A. Hossain: alamgir.hossain@anglia.ac.uk
Miltu K. Ghosh: miltu.pharm@gmail.com

*Corresponding author:

Nazia Hameed
Anglia Ruskin IT Research Institute (ARITI), Anglia Ruskin University
Room no. SAW303D Sawyers Building Bishop Hall Lane, CM1 1SQ, Chelmsford, UK
Phone: + 44 7740 311828

Abstract

Skin diseases remain a major cause of disability worldwide and contribute approximately 1.79% of the global burden of disease measured in disability-adjusted life years. Alone in the United Kingdom, 60% of the population suffer from skin disease in their life. In this paper, we proposed an intelligent digital diagnosis scheme to improve the classification accuracy of multiple diseases. In this investigation, a multi-class multi-level (MCML) classification algorithm inspired by “divide and conquer” rule is explored to address the research challenges. The MCML classification algorithm is implemented using traditional machine learning and advanced deep learning approaches. Improved techniques are proposed for noise removal in traditional machine learning approach. The proposed algorithm is evaluated on 3672 classified images, collected from different sources and the diagnostic accuracy of 96.47% is achieved. To verify the performance of the proposed algorithm, its metrics are compared with Multi-class Single-Level classification algorithm which is the main algorithm used in most of the existing literature and it is worth mentioning that the proposed algorithm outperform existing algorithms in literature. The results also indicate that MCML classification algorithm is capable of enhancing the classification performance of multiple skin lesions.

Keywords: *skin lesion classification, computer-aided diagnosis, deep learning, texture & colour features, melanoma classification, eczema classification*

1. Introduction

The human skin is the largest body organ and can be agonised from different factors like sun (Ultraviolet) radiations, tanning, lifestyle, smoking, alcohol usage, physical activities, viruses and working environment (Jaworek-korjakowska & Kleczek, 2018; Salem, Azar, & Tokajian, 2018). These factors compromise its integrity and have a profound, devastating impact on its well-being. The illness

that directly affects the skin is the fourth most frequent cause of all human diseases, affecting almost one-third of the world's population around 1.9 billion people at a time (British Association of Dermatologists, 2015), hence seeking research interest across different disciplines. Skin diseases contributed to approximately 1.79% of the global burden of diseases measured in disability-adjusted life years (DALYs) (Karimkhani, Dellavalle, & Coffeng, 2017) as graphically shown in Figure 1 .

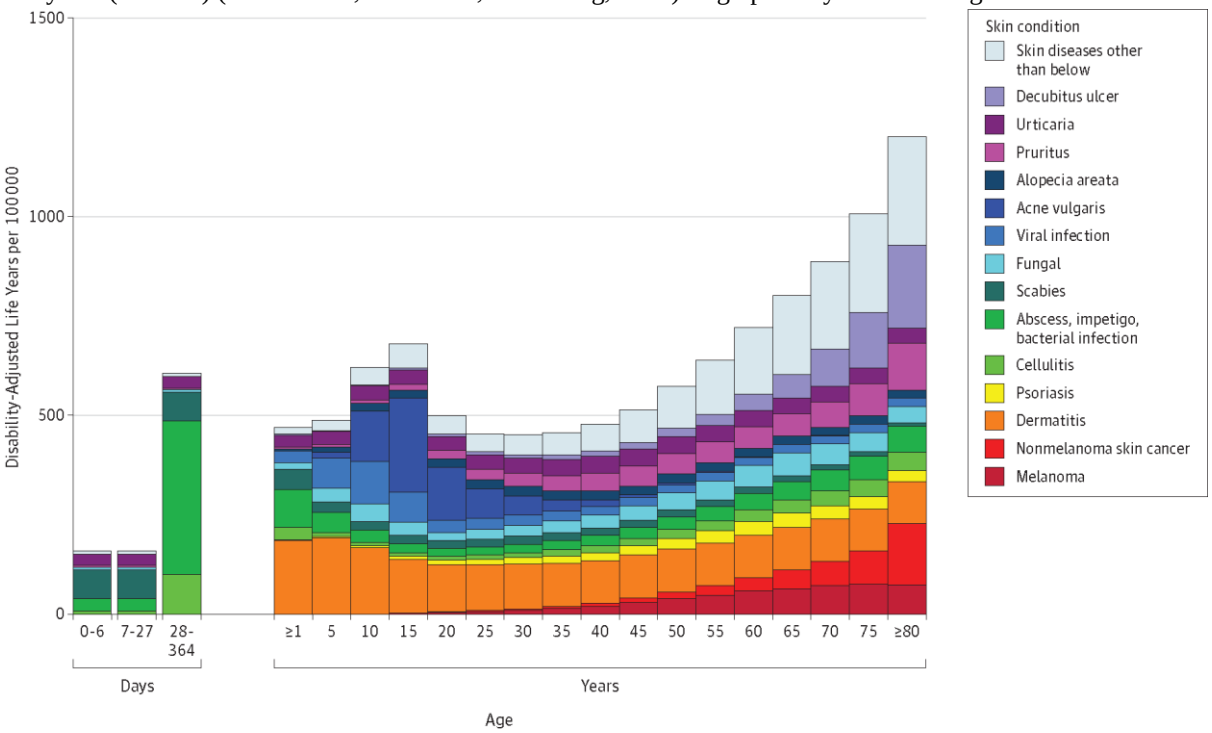

Figure 1. Skin Condition Disability-Adjusted Life Years by Patient Age (Karimkhani et al., 2017)

In the United Kingdom, 60% of the population suffer from skin diseases in their lifespan ("British Skin Foundation," 2018). Skin diseases may be cancerous, inflammatory or infectious and affect people of all ages, especially elderly and young children (British Association of Dermatologists, 2015). There are severe consequences of skin diseases like death (in case of melanoma), impairment of daily activities, loss of relationships, and damage to internal organs. Moreover, they also pose a real threat of mental illness leading to isolation, depression and even suicide (Picardi, 2013). To decrease the associated consequences, cost, mortality and morbidity rate, skin diseases should be treated in their initial stages. Cancer and eczema are among the top five common skin disorders, according to Dr Macrene Alexiades-Armenakas ("5 Most Common Skin Disorders," 2017) therefore our main focus is to develop an intelligent digital diagnosis scheme that can diagnose and classify these diseases.

Melanoma, a type of skin cancer is caused by an uncontrolled growth of melanin in the melanocytes cells. It is the most common and hastily increasing type of cancer. Melanoma is commonly classified into two types, benign and malignant melanoma (Nasir et al., 2018). In benign lesions (common nevi) melanin is normally present in the epidermis layer. Melanin is reproduced at a high abnormal stage in the malignant lesions as shown in Figure 2. Malignant lesions are not life-threatening till the melanocytes, and their associated cells remain in the epidermis layer but when they penetrate in the dermis and leave their deposits then the nature of the skin colour changes, and it became dangerous (Hameed, Hassan, & Hossain, 2016; Maglogiannis & Doukas, 2009). According to World Health Organization, between 2 and 3 million non-melanoma skin cancers and 132,000 melanoma skin cancers occur globally each year (World Health Organization, 2018).

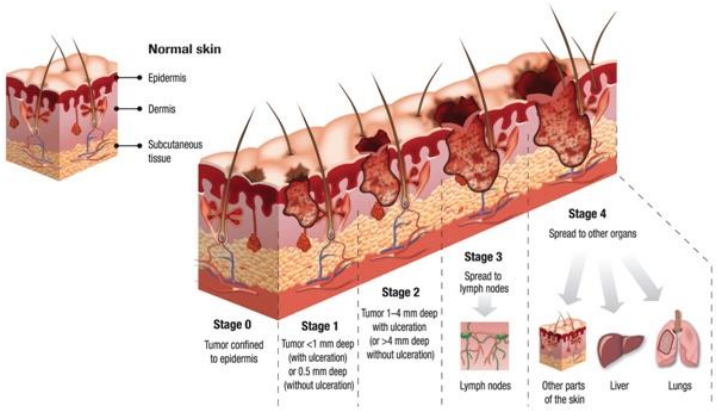

Figure 2. Spreading of melanoma in Epidermis and Dermis Layer ("Melanoma - Skin Dermatologists," 2016)

Approximately 99,550 cases are diagnosed in the USA, and approximately 13,460 are fatal (Siegel, Miller, & Jemal, 2018). In spite of all these facts, melanoma is the most treatable cancer if detected at early stages. If skin cancer is detected in stage 1, the survival rate is almost 96%, whereas it is decreased to only 5%, if detected at stage-IV (Freedberg, Geller, Miller, Lew, & Koh, 1999). Due to its life-threatening nature, it has gained remarkable attention from research and healthcare community, and their ultimate goal is to diagnose it in early stages. However, it is challenging due to similarities in melanocytic and non-melanocytic skin lesions. Eczema is an inflammatory disease and caused by many factors. In the literature, most of the work is done on skin cancer classification, and limited work is available on the classification of other diseases.

In this research work, a multi-class multi-level (MCML) algorithm is proposed and developed to provide multi-class classification of skin diseases. In MCML, the skin lesion classification problem is divided into sub-problems, and these sub-problems are solved in multiple steps instead of only utilising one step to perform the classification. The MCML algorithm is implemented using two techniques: traditional machine learning approach and deep learning approach. In the traditional machine learning technique, improved techniques are proposed for removing black frames and circles which is another contribution of this research work. Set of features that can be extracted from every disease is also listed. To demonstrate the improved performance of the proposed algorithm, results are compared with the multi-class single-level algorithm as well as with existing research work. Comparison of traditional machine learning and deep leraning approach is also performed in this research work.

The rest of the paper is organized as follows. Section 2 reviews the existing literature. Material are given in section 3, while methods are discussed in section 4. Experiments and results are explained in section 5. Section 6 conclude the research and highlight future directions.

2. Related Work

Over the last two decades, researchers have worked to provide the intelligent diagnosis systems for the automated classification of skin disorders to assist the dermatologists, primarily in early classification of skin cancer (E.Umbaugh, H.Moss, & V.Stoecker, 1992; Ercal, Moganti, Stoecker, & Moss, 1993; Nischik & Forster, 1997; Zhang, Stoecker, & Moss, 2000; Vasconcelos & Vasconcelos, 2017; Dorj, Lee, Choi, & Lee, 2018;). With the advancement of computer vision and image processing, continuous improvement is required to provide better accuracy. In the literature, intelligent diagnosis systems have been developed using traditional machine learning approach (De Guzman, Maglaque, Torres, Zapido, & Cordel, 2015; De Guzman et al., 2015; Alam et al., 2016; Oliveira, Pereira, Manuel, & Tavares,

2016; Zakeri & Hokmabadi, 2018) and advanced deep learning approach in the few recent years (Esteva et al., 2017; Vasconcelos & Vasconcelos, 2017).

Intelligent diagnosis systems based on traditional machine learning techniques mostly consist of pre-processing, segmentation, feature extraction and classification phases. In pre-processing, images are pre-processed to remove the noise and to improve the segmentation accuracy (Oliveira, Marranghello, Pereira, & Tavares, 2016). Noise can become part of images due to many factors such as capturing environment, capturing device, and lightening condition which may affect the images in the form of a black frame, dermoscopic gel, air bubbles, circles, skin lines, hairs, blood vessels etc (Hameed et al., 2016; Maglogiannis & Doukas, 2009). Pre-defined masks are used by Sultana et al. to remove the black frames (Sultana, Dumitrache, Vocurek, & Ciuc, 2014). Similar work is done by Abuzaghleh et al. to remove the black frame (Abuzaghleh, Barkana, & Faezipour, 2014). However, these techniques can only work when the dataset is small and consistent. To remove the hairs, DullRazor (Lee, Ng, Gallagher, Coldman, & McLean, 1997) is the widely used technique in literature. Several approaches have been used in the literature for noise removal and enhance the quality of images include image resizing (Jain, Jagtap, & Pise, 2015), contrast adjustment, filtering (Abbas, Celebi, & Fondón, 2011; Celebi, Aslandogan, & Bergstresser, 2005; Maglogiannis, Ieee, & Delibasis, 2015), cropping and colour quantization (Celebi et al., 2005). Segmentation is the following step of the noise removal step to extract the region of interest (ROI). Image segmentation techniques have been developed based on several techniques such as thresholding, clustering, region based, and soft computing to get the ROI (Oliveira, Pereira, Manuel, & Tavares, 2016). Among these techniques, thresholding is mostly used because of its simple nature. Clustering such as k-means is also used, but it requires identifing number of K before applying it. However, a single technique may not work, and in turn a hybrid techniqe works better when dealing with images of diverse nature. After identifying the ROI, a number of features are extracted from it to help perform the final classification. Geometric features (Oliveira, Pereira, & Tavares, 2018), colour features (Nasir et al., 2018; Oliveira et al., 2018) and texture features (Oliveira et al., 2018; Zakeri & Hokmabadi, 2018) are widely used for performing the classification. However, all these features cannot be extracted from every skin lesion because of different factors like disease nature, presence of moles, area effected by disease, capturing distance etc. (Salem et al., 2018). For the classification process, several machine learning classifiers have been applied to the extracted features to achieve the best results. Frequently used techniques for skin lesion classification are support vector machine (Alam et al., 2016; C.-Y. Chang & Liao, 2011; Hameed et al., 2016), artificial neural networks (ANN) (Rubegni et al., 2002; Shamsul Arifin, Golam Kibria, Firoze, Ashraful Amini, & Yan, 2012), k-nearest neighbours (Çataloluk & Kesler, 2012; Ganster et al., 2001) and decision trees (Salem et al., 2018; Victor & Ghalib, 2017).

Using traditional machine learning approach, most of the research is done on the classification of skin cancer and limited research is performed on the classification of other diseases. Out of this limited research, some intelligent diagnosis systems are trained on the clinically extracted features instead of images (C. L. Chang & Chen, 2009; Güvenir & Emeksiz, 2000; Kumar, Kumar, & Saboo, 2016; Übeyli, 2009; Xie & Wang, 2011). Erythemato-squamous diseases consisting psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis, and pityriasis rubra pilaris are classified by the Derya et al. using SVM and neural network with an accuracy of 98.32% and 97.7% respectively (Übeyli, 2008, 2009). Similar kind of work is done by Guvenier and Emeksiz where they presented an expert system for classification of Erythemato-squamous diseases by incorporating nearest neighbour, naïve bayesian and voting feature algorithm. Voting feature algorithm outperforms with an accuracy of 99.2% (Güvenir & Emeksiz, 2000). Chang used a hybrid technique comprising the features of the neural network and decision tree to construct a predictive model for diagnosing Erythemato-squamous diseases using multi-variate variables. Their proposed predictive model achieves an accuracy of 92.62% (C. L. Chang & Chen, 2009). Abdi and Giveki proposed automated detection of erythemato-squamous diseases using PSO–SVM based on association rules (Abdi & Giveki, 2013). Xie et al., Kumar et al. and Nanni et al. also classify the erythemato-squamous disease using machine learning

technique with an accuracy of 98.61%, 97.22% and 95% respectively (Kumar et al., 2016; Nanni, 2006; Xie & Wang, 2011). All of the work done for classification of Erythemato-squamous diseases is performed on clinically extracted multivariate features after biopsy (Ilter & Guvenir, 1998). Disease classification using clinically extracted features is not feasible as it is time consuming and difficult to achieve. Moreover, it requires domain expertise and expert knowledge.

Other limitations found in literature are single disease classification and training on limited data (Ganster et al., 2001; Nugroho et al., 2013; Giotis et al., 2015; Jain et al., 2015; Alam et al., 2016;). Continuing with the single disease classification, Dorj et al. used the SVM classifier on the features obtained by CNN to classify the skin cancer. They trained and tested their algorithm on 3753 images which were collected from the internet and achieved an accuracy of 94.2% (Dorj, Lee, Choi, & Lee, 2018). Nasir et al. presented a strategy for the classification of melanoma with an accuracy of 97.5% when tested on the PH^2 dataset (Nasir et al., 2018). Although they have achieved good classification accuracy, their trained model lacks adaptability to new data because of the training on limited data. Zakeri et al. proposed a hybrid classifier for detecting the cancerous lesions with an accuracy of 96.8%, 97.3% and 98.8% for the melanoma, dysplastic and benign on 792 images. Work done by Esteva et al. achieves dermatological level classification using convolutional neural network (CNN); a deep learning approach but they also worked only on the skin cancer. Vasconcelos et al. also performed experiments with different variations using deep learning for the melanoma image analysis. All the above-mentioned research work is performed for classifying a single disease. Alam et al. presented a model for classification of healthy and eczema images. An accuracy of 90% is obtained when they trained and tested their model on 85 images. Their model also suffers from the generalisation because of limited data. The comparison of skin diseases classification work available in the literature is given in Table 1 where limitations are highlighted.

Table 1: Summary of studies in literature for skin disease classification

Reference	Classification Categories	Images	Results
Salem et al., 2018	**Melanoma**	369	Acc*: 76.17%
Dorj et al., 2018	**Melanoma**	3753	Acc: 94.2%
Oliveira et al., 2018	**Melanoma**	1104	Acc: 92.3%
Nasir et al., 2018	**Melanoma**	200	Acc: 97.7%
Alam et al., 2016	Healthy & Eczema	**85**	Acc: 90%
Übeyli, 2008, 2009	Erythemato-squamous Diseases[1]	**Non-image data**	Acc: 98.32%
Güvenir & Emeksiz, 2000	Erythemato-squamous Diseases	**Non-image data**	Acc: 99.2%
C. L. Chang & Chen, 2009	Erythemato-squamous Diseases	**Non-image data**	Acc: 92.62%

[1] Erythemato-squamous Diseases includes psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis, and pityriasis rubra pilaris

Abdi & Giveki, 2013	Erythemato-squamous Diseases	**Non-image data**	Acc: 98.91
Kumar et al., 2016	Erythemato-squamous Diseases	**Non-image data**	Acc: 98.61%

Acc* = Accuracy

3. Dataset

To conduct this research, a combination of the images from different sources are used. Sources for collecting dataset include open-access dermatology repositories, organizations and researchers. Open-access dermatology repositories include the International Skin Imaging Collaboration (ISIC) Dermoscopic Archive (Codella et al., 2016) and the PH[2] (Mendonca, Ferreira, Marques, Marcal, & Rozeira, 2013). The ISIC melanoma project is an industry and academia partnership designed to facilitate the application of digital skin imaging to help reduce melanoma mortality. The ISIC Archive dataset[2] constitutes 13000 melanocytic lesions images that are biopsy proven and annotated as either benign or malignant. The PH[2] dataset[3] is also a publically available database comprises 200 images which include the manual segmentation along with the clinical diagnosis performed by expert dermatologists. For healthy dataset, 11K hands dataset is used (Afifi, 2017). Apart from 11k, some of the heathy images are collected by the authors themselves. Melanoma and eczema images are also obtained from different organizations like DermIS (DermIS, 2018), DermQuest (DermQuest Image Library, 2018) and DermNZ ("DermNZ," 2018). These organizations provide classified images of different skin lesions, and they are freely available to use for the academic purpose. A subset of the dataset is also obtained from Alam et al. (Alam et al., 2016). This dataset contains 85 images belonging to healthy and eczema category. Some of the images along with their categories are shown in Figure 3.

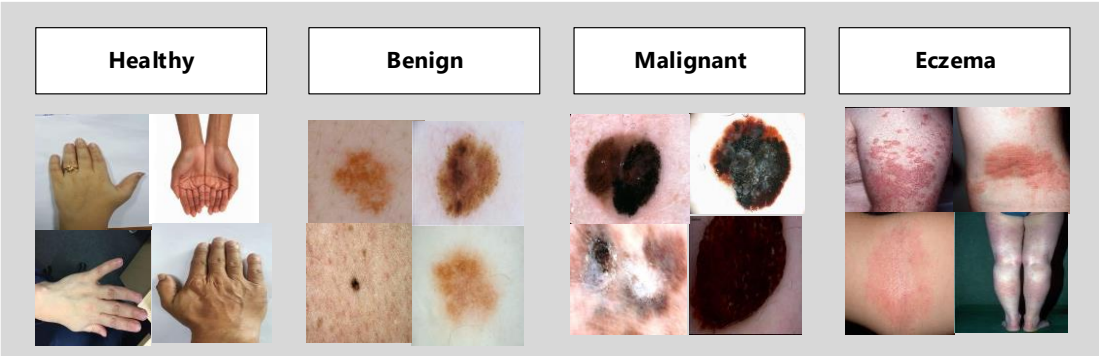

Figure 3. Example images from the dataset belonging to different categories

The number of images in each category are not consistent, arising data imbalencing problem. Data imbalencing issue is very critical, as it may affect the classification results (Japkowicz & Stephen, 2002). Different approaches are proposed in the literature (Burdick, Marques, Weinthal, & Furht, 2017; Japkowicz & Stephen, 2002) to address this issue. In this research work, we are using down sampling approach. Eczema has the minimum number of images, so other categories are downsized to the number of images in eczema category. The total number of images in each category are downsized randomly to 918 to make them equivalent. Images in each category are divided into two parts, one is

[2] The ISIC dataset is available at: https://isic-archive.com/#images
[3] The PH[2] dataset is publically available at : https://www.fc.up.pt/addi/ph2%20database.html

for training and testing and other is for validation. For training and testing, each category has 860 images whereas 58 images are for cross-validation for each category.

4. Methodology

Two approaches are used in this research work for classifying skin lesions. The first approach makes use of traditional machine learning while deep learning is used in the second approach. Each approach is explained in the next subsections.

4.1. Traditional Machine Learning approach

In this paper, we define traditional machine learning approach as a computational approach for skin lesion classification that learn from a predefined bag of features and go through different steps. These steps are illustrated in Figure 4 which involves the following: (1) pre-processing, (2) segmentation, (3) feature extraction and (4) classification. Each step is detailed below.

4.1.1. Pre-processing

Main purpose of the pre-processing step is to remove noise from the image. While capturing the image, there are many variables that can affect the image such as skin nature, capturing environment, capturing device and lightening condition. Due to these variables; images may contain some artefacts such as black frames, dermoscopic gel, air bubbles, skin lines, hairs, and blood vessels. These artefacts create a barrier to the segmentation process and result in accuracy loss and increased computational cost (Hameed et al., 2016; Maglogiannis & Doukas, 2009). Therefore, some pre-processing steps are required to remove these artefacts and in turn facilitate the segmentation and classification process. The tasks perform in the pre-processing step are shown in Figure 5.

4.1.1.1. Image Resizing

The data used in this research work is gathered from different sources hence the images are of different dimensions. For consistency, all the images are resized to make the images consistent. In the image resizing step, all the images are resized to 720 x 720 pixels.

4.1.1.2. Hair removal

Hair is an important noise as they can degrade the system performance and create an obstacle in achieving good accuracy in the segmentation hence in the classification step (George, Aldeen, & Garnavi, 2015; Maglogiannis et al., 2015; Toossi et al., 2013). Therefore, to achieve better segmentation, hair should be removed. For hair removal, an image processing based technique is implemented, and the algorithm is given in Algorithm 1.

Algorithm 1. Hair removal algorithm

Input: Image with hair ($I_{RGBwithHairs}$)

Output: Hair free image

Initialization:

$\mathbf{I_{RGBwithHairs}}$ ← Input RGB image with hairs

$\mathbf{I_{Grey}}$ ← Convert $\mathbf{I_{RGBwithHairs}}$ to grey image

$\mathbf{I_{Hist}}$ ← Apply Histogram equalisation on $\mathbf{I_{Grey}}$

$\mathbf{I_{Average}}$ ← Apply the mean filter of window size 9x9 on $\mathbf{I_{Hist}}$

$\mathbf{I_{Mask}}$ ← $\mathbf{I_{Grey}} - \mathbf{I_{Average}}$

$\mathbf{I}_{\mathbf{Thresh}}$	← Apply thresholding on $\mathbf{I}_{\mathbf{Mask}}$
$\mathbf{I}_{\mathbf{Binary}}$	← Convert $\mathbf{I}_{\mathbf{Thresh}}$ into a binary image
$\mathbf{I}_{\mathbf{Morp}}$	← Remove unnecessary details from $\mathbf{I}_{\mathbf{Binary}}$
Mask	← Convert black pixels to white and vice versa
$\mathbf{Mask}_{\mathbf{BI}}$	← Apply bilinear interpolation on **Mask** and hair are replaced with neighbour pixels
$\mathbf{I}_{\mathbf{RGBwithoutHairs}}$	← Apply $\mathbf{Mask}_{\mathbf{BI}}$ on $\mathbf{I}_{\mathbf{RGBwithHairs}}$

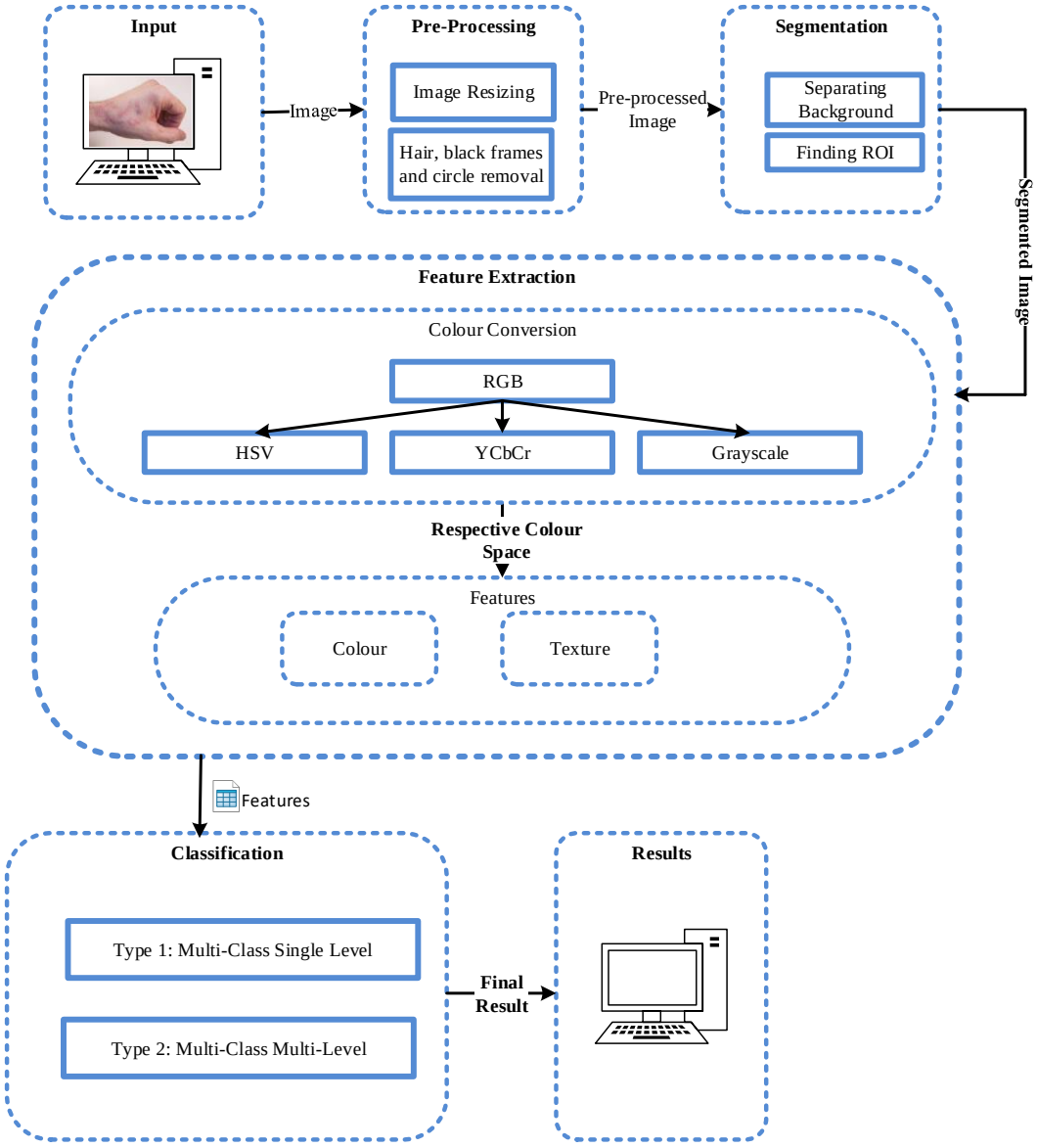

Figure 4: Overall methodology to classify skin disease using traditional machine learning process

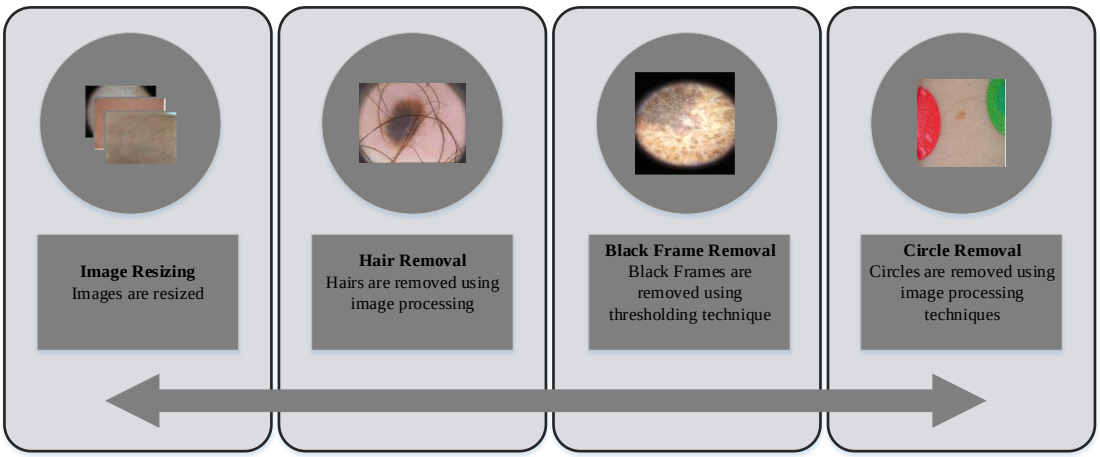

Figure 5: Pre-processing Step

In the first step, the RGB image is given as an input. The RGB image is converted to a grey image and histogram equalisation is applied to adjust the values of pixels. To further smoothen the image, mean filter is applied. Experiments are performed with 3x3, 5x5, 7x7 and 9x9 window size. The window size filter 9x9 is selected because it gives better results. To highlight the hairs, hair mask is created by subtracting the grey image and the average image. To further improve the mask thresholding is applied on the mask and then converted into a binary image. There are unnecessary artefacts in the binary image which are not the hair particles; they should be removed. For removing these particles, morphological erosion operation is applied. The mask is complemented, and bilinear interpolation (Mathworks, 2018) is applied to the mask, and the hair pixels are replaced by their neighbouring pixels. Results of each step of hair removal algorithm are graphically shown in Figure 6. Final mask is applied to the original image to get the hair-free image.

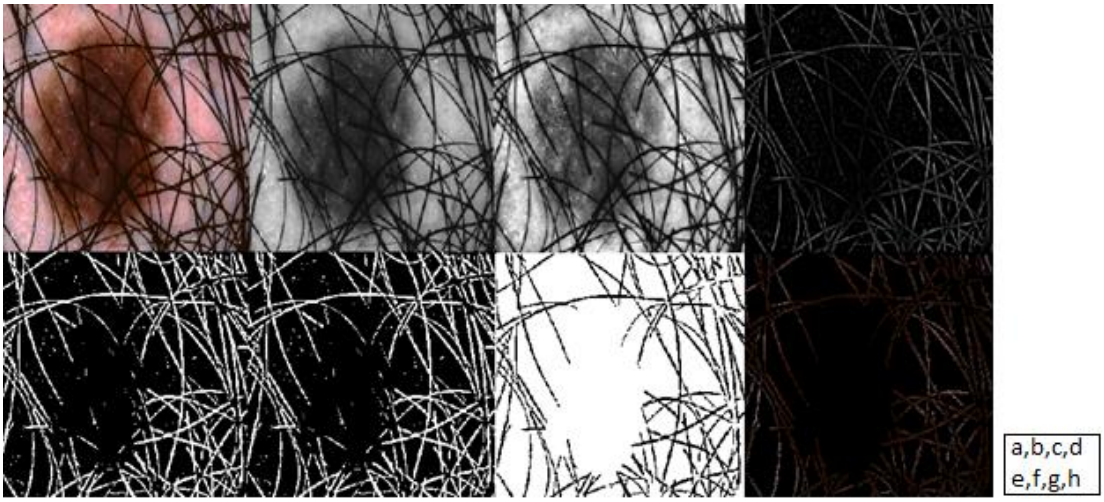

Figure 6. Steps involved in the removal process a) RGB image b) Gray image c) Contrast-enhanced image d) Mask Image e) Binary Image f) Morphological image g) Hair Mask f) Mask after bilinear interpolation

4.1.1.3. Black Frame Removal

The black frame is another noise that creates another obstacle in the segmentation phase. In the literature, there are some techniques for removing the black frames by creating a mask of different shapes (Abuzaghleh et al., 2014; Sultana et al., 2014). These techniques can work for few images because one can create masks for few images but when you have thousands of

images, this technique will not work. In this research work, an enhanced technique is proposed to dynamically generate a black frame mask at runtime. Afterwards, this mask is used to get the image without black frame. The proposed technique is a pixel-based technique and works on the RGB images. In the proposed technique, first, the red, green and blue channels of RGB colour space is extracted, and their mean is calculated, and thresholding technique is applied to calculate the black mask. Once the mask is calculated, the final mask is used to get an image without black frame. Figure 7 shows the flowchart of the black frame removal process.

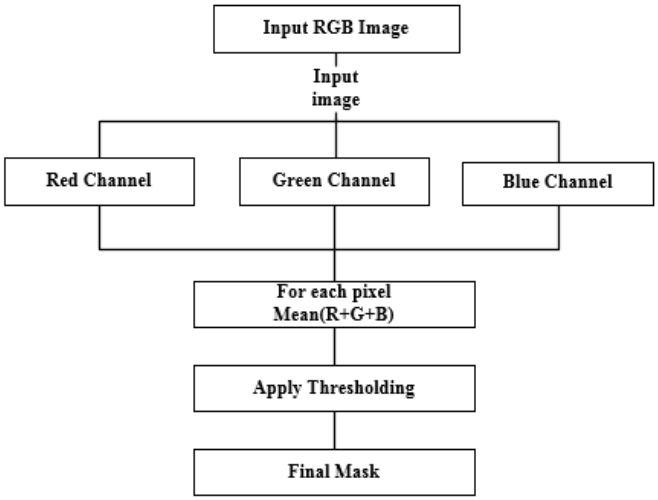

Figure 7. Black frame removal flowchrat

In our experiments, we have tested different threshold value and obtained the best results by setting the threshold value to 50. The proposed technique is applied to the dataset collected from different sources, and the black frame removal success rate is 99%. Some of the results are shown in Figure 8. which shows that the proposed technique provides promising results on the images containing a variety of black frames.

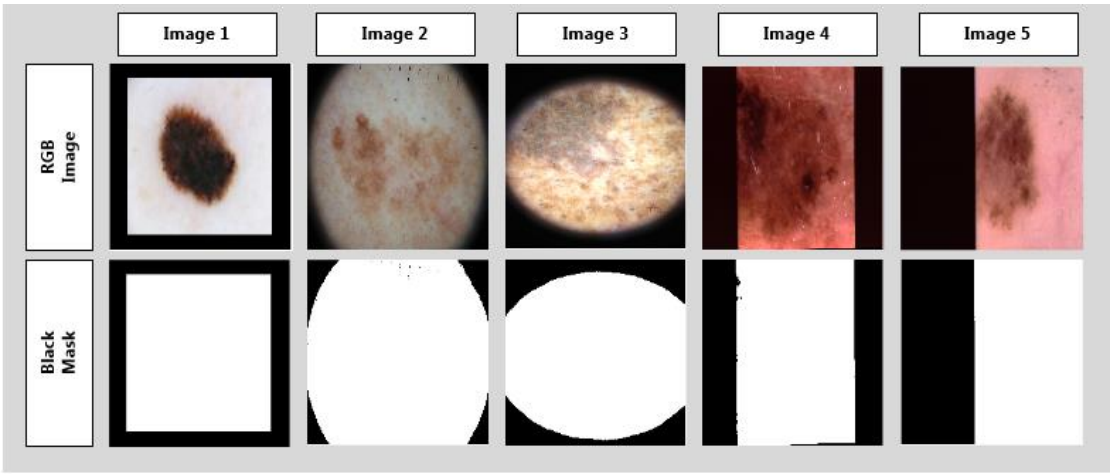

Figure 8. Experimental results of the black frame removal process

The proposed technique outperforms techniques proposed by Abuzagleh et al. and Sultana et. al (Sultana et al., 2014;Abuzaghleh, Barkana, & Faezipour, 2014) as the mask is dynamically calculated according to the input image at the runtime instead of a pre-defined mask.

4.1.1.4. Circle Removal

Some images in ISIC challenge dataset contains the circle of varied sizes and creates difficulty in the segmentation process as shown in Figure 9.

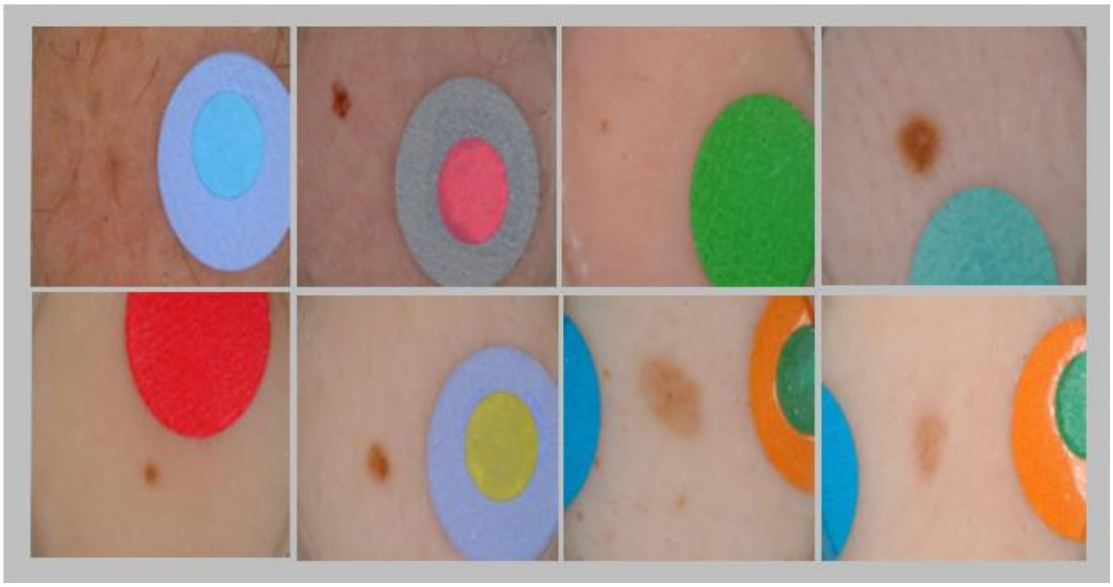

Figure 9. Images with circle noise

For removing the circles, the input image is converted into a grey image. After converting to the grey image, a binary image is calculated by applying the Otsu's thresholding technique. Otsu's thresholding method computes a global threshold level, which is used to convert a grey image to a binary image. The obtained binary image after thresholding contains unnecessary details which are removed by applying morphological erosion operation. For morphological erosion process, the "disk" structuring element of size one is used in this algorithm. After erosion operations, hole filling is applied to get the final mask. The algorithm of the circle removal process is illustrated in Algorithm 2 while the results are presented in Figure 10.

Algorithm 2. Skin images containing circles of different sizes

Input: Image with circles ($I_{RGBwithCircle}$)

Output: $I_{RGBwithoutCircle}$

Initialization:

$I_{RGBwithCircle}$ ← Input RGB image with circles

$\mathbf{I_{Grey}}$ ← Convert $\mathbf{I_{RGBwithCircle}}$ to grey image

$\mathbf{I_{Thresh}}$ ← Apply thresholding on $\mathbf{I_{Grey}}$

$\mathbf{I_{Binary}}$ ← Convert $\mathbf{I_{Thresh}}$ into a binary image

$\mathbf{I_{Erosion}}$ ← Remove unnecessary details from $\mathbf{I_{Binary}}$ by applying erosion operation

$\mathbf{I_{Mask}}$ ← Fill the holes in $\mathbf{I_{Erosion}}$

$I_{RGBwithoutCircle}$ ← Apply $\mathbf{I_{Mask}}$ to $I_{RGBwithCircle}$

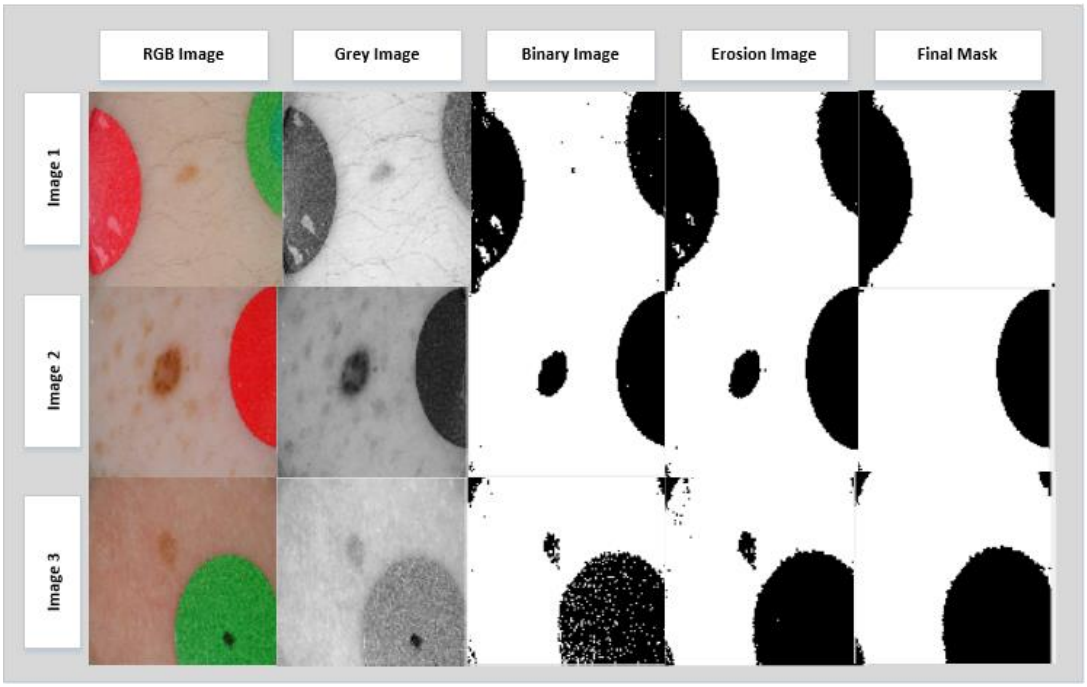

Figure 10. Steps involved in removing circles from skin images

4.1.2. Segmentation

The main purpose of a segmentation step is to get the ROI because ROI is expected to have more relevant information in the form of different features that can be used for lesion classification and diagnosis. The discontinuity and similarity of some properties of ROI to be segmented is the basis of the segmentation process. For extracting the ROI, a hybrid technique is used. This hybrid technique utilizes the capability of K-means clustering, thresholding and morphological operations. In thresholding, Otsu's method is applied. For this, the image is first converted into grayscale and grey thresh is computed. This grey thresh is applied on the image to obtain the region of interest. As we need the RGB image, the grey image is converted back to the RGB Image. K-means clustering with k=2 is applied, and the cluster with ROI is selected. The final segmented image is obtained by combining the Otsu's and a clustering result. Finally, unnecessary details are removed from the segmented image by applying the morphological operations. The flowchart of the segmentation process is shown in Figure 11.

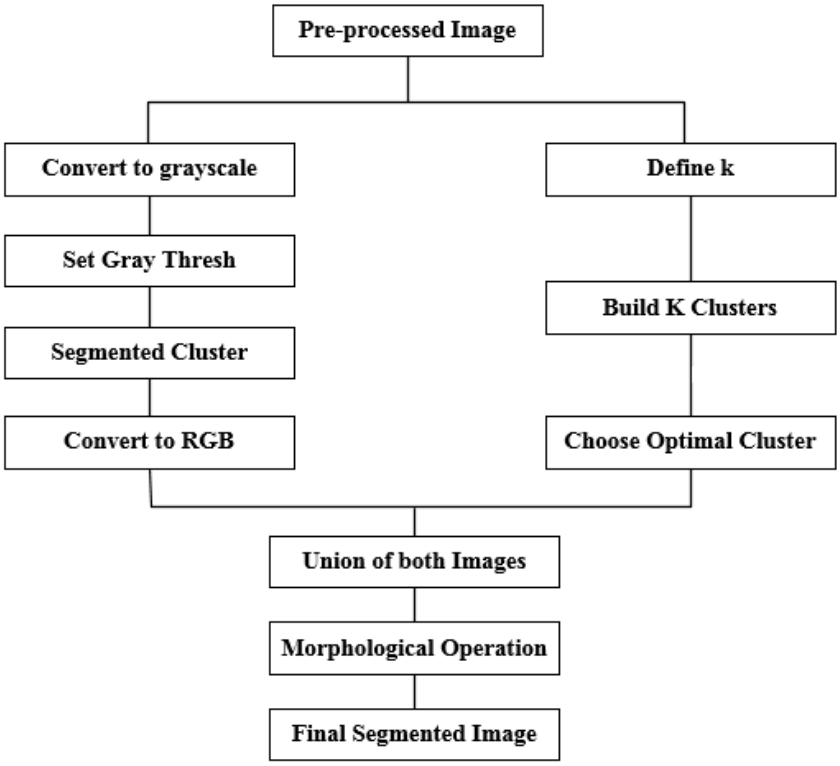

Figure 11. Steps involved in the segmentation process

In some cases, the Otsu's thresholding technique alone is sufficient to extract the ROI. However, in the majority of the cases, the hybrid technique gives better results and the extracted ROI is more accurate. Results of the segmentation technique are illustrated in Figure 12.

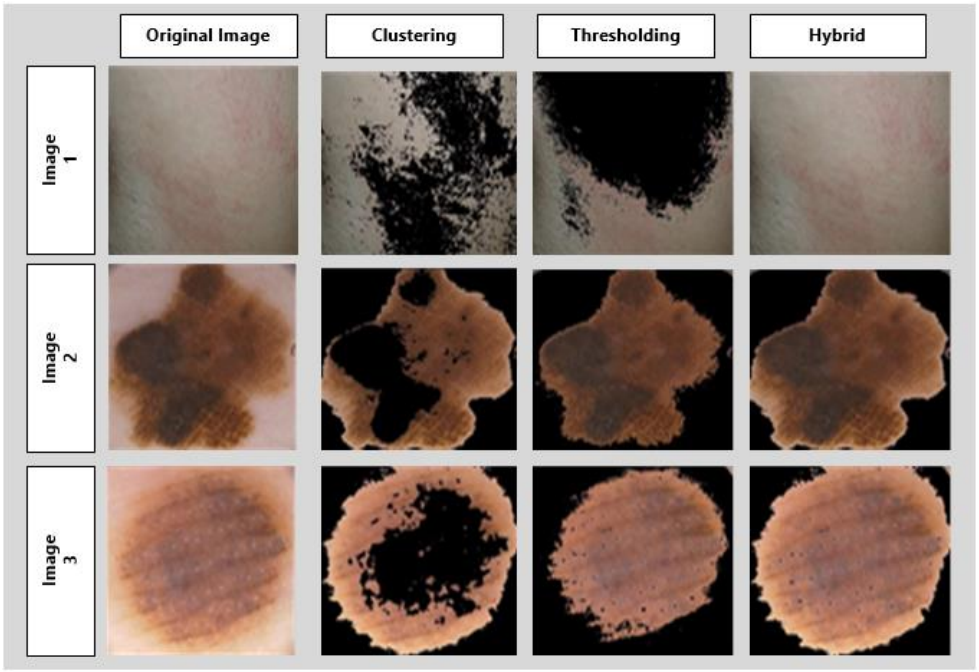

Figure 12. Image segmentation results obtained after applying clustering, thresholding and hybrid technique

4.1.3. Feature Extraction

One of the main challenges in this research work is to find the set of features that can be applied to all images, i.e. healthy, benign, malignant, and eczema, as the images of each category differ from others and they have their own characteristics. For example, mostly the benign and malignant images have clear boundary and one can extract border and shape features while the healthy and eczema images may cover the full body part and don't have clear boundary and it's difficult to extract the border features. Thus, consistent set of features are required that can be extracted from images of all categories. In feature extraction, 36 features belonging to colour and texture categories are extracted and stored in the feature vector for the classification step.

From the segmented image, the colour histogram of R, G and B channel is calculated, and features are extracted from it. Colour histogram is a representation of the number of pixels at each intensity level of a colour channel. Colour histogram features state the global properties of the intensity level distribution for each colour channel. The histogram of an image can be represented as a probability distribution, *P(g),* of the intensity levels as given in (1).

$$P(g) = \frac{N(g)}{M} \dots\dots (1)$$

Where g is the intensity level, *N (g)* is the number of pixels at intensity level g, and M is the total number of pixels. The histogram features extracted from each colour channel are mean, mode, standard deviation, skewness, energy, entropy and kurtosis. The mean ($\bar{g}$) reflects the overall intensity level in the image. Standard deviation; also known as the variance and gives the spread of the data. A high contrast image has a high variance and a low contrast image has a low variance. The mathematical representation of mean and standard deviation is given in (2) and (3).

$$M(\bar{g}) = \sum_{g=0}^{W-1} gP(g) = \sum_r \sum_c \frac{I(r,c)}{M} .. (2)$$

$$SD(\sigma_g) = \sqrt{\sum_{g=0}^{W-1} (g - \bar{g})^2 P(g)} \dots (3)$$

where W is the number of intensity levels, r is the number of rows and c is the number of columns in the image. The skewness computes the asymmetry of the probability distribution of the histogram. Therefore, it reveals information about the shape of the distribution. Skewness of image is calculated using (4).

$$Skewness = \frac{1}{\sigma^3} \sum_{g=0}^{W-1} (g - \bar{g})^3 P(g) \dots\dots (4)$$

The energy measure is related to the colour span (i.e. the spread of the pixel values). The pixel colour energy decreases as the pixel values span a wider intensity range. The entropy measure describes the required amount of information to code the image data. In contrast to the energy measure, the entropy increases as the pixel values span a wider intensity range. The mathematical representation of energy and entropy is given in (5) and (6).

$$Energy = \sum_{g=0}^{W-1} [P(g)]^2 \dots\dots (5)$$

$$Entropy = -\sum_{g=0}^{w-1} [P(g)\, log_2\, P(g)] \dots\dots (6)$$

For extracting colour features, different colour spaces are used. Minimum and maximum values are extracted from each channel in the RGB colour space which are denoted as R_{min}, R_{max}, G_{min}, G_{max}, B_{min}, B_{max} for red, green, and blue channels respectively. For extracting other colour features the RGB

image is divided into HSV, YCbCr, and grey scale colour spaces and consequently different features are extracted. These features include mean value of H channel, V channel, Cb channel, Cr channel and gray image which are denoted as H_{mean}, V_{mean}, Cb_{mean}, Cr_{mean} and $Gray_{mean}$ respectively thus making 11 colour features in total.

For extracting texture features, Gray Level Co-occurrence Matrix (GLCM) is computed first, and then features are computed from it. GLCM is a matrix which shows the distribution of co-occurring pixels values at a given offset. The GLCM functions are used to characterize the texture of image by calculation the occurrence of pair pixels, and their spatial relationship occurs in the images and then calculating the statistical measurements (MATLAB, 2017). The features extracted from this category are contrast, correlation, energy, and homogeneity. The contrast measures the local variations in the GLCM. To measure the joint probability occurrence of specified pair pixels, correlation is used. Contrast and correlation of GLCM are computed using mathematical equation given in (7) and (8)

$$Contrast_{GLCM} = \sum_{i,j=0}^{W-1} P_{ij}(i-j)^2 \dots (7)$$

$$Correlation_{GLCM} = \sum_{i,j=0}^{W-1} P_{ij} \frac{(i-\mu)(j-\mu)}{\sigma^2} \dots (8)$$

where P_{ij} is the element of GLCM, μ is the GLCM mean and σ is the variance of the intensities of all reference pixels in the relationships that contributed to the GLCM. Energy is also known as the uniformity or the angular second moment, and it provides the sum of squared elements in the GLCM. The energy is calculated using the same formula given in (5), but here the input is the GLCM matrix. Homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. The mathematical formula to calculate the homogeneity is given in (9). A total of 36 features are extracted and stored in a vector to be used in the classification step.

$$Homogeneity_{GLCM} = \sum_{i,j=0}^{W-1} \frac{P_{ij}}{1+(i-j)^2} \dots (9)$$

4.1.4. Classification

To classify an image using extracted features, an appropriate learning algorithm should be selected. There are two types of classification algorithms available in machine learning, i.e. supervised learning algorithms and unsupervised learning algorithms. Supervised learning algorithms take the classified input data and train a model to generate the predictions for the response to the new unseen data. Unsupervised algorithms are used when classified data is not available. For this research, supervised learning algorithms are suitable because of the availability of the classified data. For the classification task, we use ANN with back propagation; a supervised machine learning technique. In training the model, 10 hidden layers is used and sigmoid as an activation function is also used to train the model. The general layout of the ANN is shown in Figure 13.

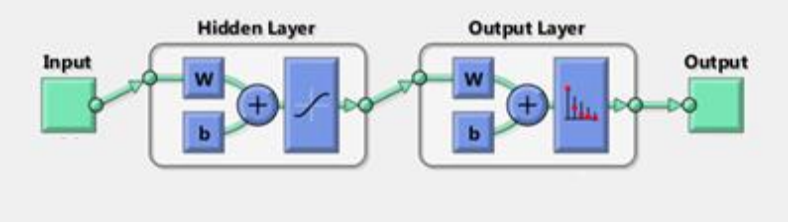

Figure 13. Artificial Neural Network Framework

For validation and evaluation of the output model (i.e. classifier) hold out technique is used (Zheng, 2015). The holdout validation is the simplest validation technique in which the dataset is divided into two sets: the training set and testing set. The training set is used for training the classifier, and the

testing set is used to test the accuracy of the classifier. The classifier learns the weights using the training set. Then these learned weights are applied to the unseen set: test data, and in turn the accuracy is measured. The results depend on the data in the training and testing dataset, and the final evaluation may be significantly different depending on the data division. The main advantage of this method is that it takes less time for computation. MCML and MCSL algorithms are applied. In MCSL classification, all the skin diseases are categorized at the same level, i.e. healthy, benign, malignant, and eczema. The feature vector extracted in feature extraction phase are given as input to the machine learning classifier. The MCSL classification is graphically shown in Figure 14.

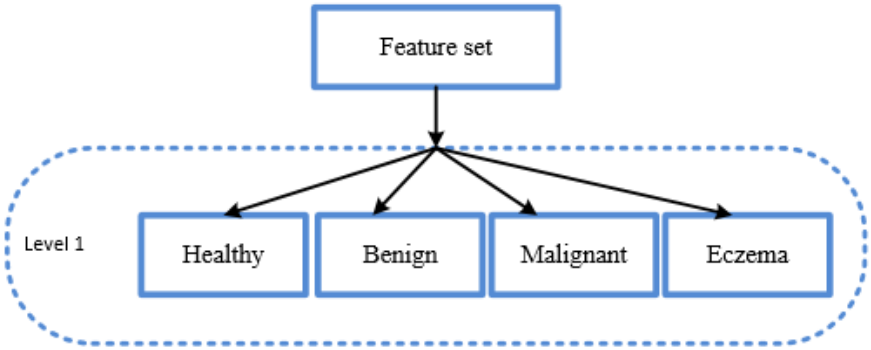

Figure 14. Multi-Class Single Level Classification

MCML works on the divide and conquer rule. In MCML, the classification problem is divided into sub-classification problems and at the end combine them to get the final result. MCML is graphically shown in Figure 15.

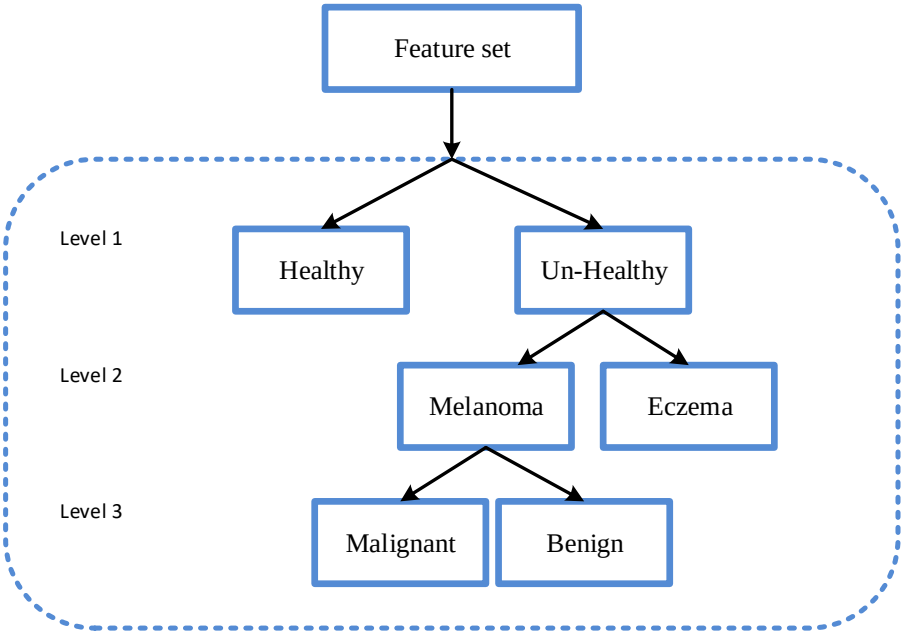

Figure 15. Multi-Class Multi-Level Classification

At level 1, binary classification is performed, and the images are classified into two categories, i.e. healthy and unhealthy. At level 2, unhealthy images are further classified into melanoma and eczema. Level 3 is the last level of MCML classifier, and at this level, the melanoma images are classified as either malignant or benign.

4.2. Deep Learning Approach

Deep learning approach is powered by the advances in computation and has been shown exceptional performance in object recognition and classification (Burdick et al., 2017; Esteva et al., 2017). Deep learning has produced results comparable to and in some cases superior to human experts. The deep learning algorithm is trained end-to-end directly from raw image pixels and the image label. For deep learning convolutional neural network (CNN) is used for image classification (Krizhevsky, Sutskever, & Hinton, 2012). For performing MCSL and MCML classification using deep learning approach, author's employs transfer learning approach, and a pre-trained model AlexNet (Krizhevsky et al., 2012) is modified, fine-tuned and re-trained on our own dataset. The AlexNet model offer several benefits like well-known implementation, few training parameters and extreme validity (Dorj et al., 2018). The deep learning layout is graphically shown in Figure 16.

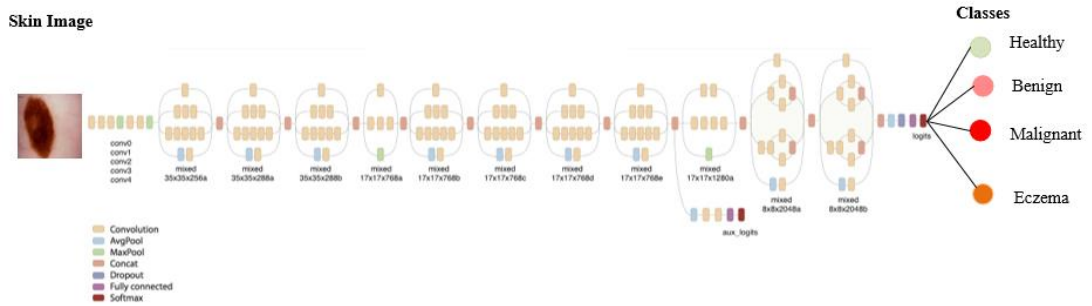

Figure 16. Convolutional neural network architecture

AlexNet is a pre-trained CNN, trained on 1,000 object classes with a top-five error rate of 15.3%. For training, the images are resized to 277 x 277 because of model constraint. The parameters used in training our model using deep learning approach are given in Table 2.

Table 2. Parameters used in training the deep learning algorithm

Parameter Name	Value
Learning Rate	0.0001
Max Epoch	20
Mini Batch Size	64

5. Results and Discussion

After performing the classification task, results of MCML and MCSL algorithms are calculated and compared using accuracy, sensitivity, specificity, and precision metrics. The accuracy, precision, sensitivity, and specificity are calculated using equations 10-13.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \text{x}100 \ldots\ldots\ldots (10)$$

$$Precion = \frac{TP}{TP + FP} \text{x}100 \ldots\ldots\ldots (11)$$

$$Sensitivity = \frac{TP}{TP + FN} \text{x}100 \ldots\ldots\ldots (12)$$

$$Specificity = \frac{TN}{TN + FP} \text{x}100 \ldots\ldots\ldots (13)$$

The terms true positive, true negative, false positive and false negative are explained in Table 3.

Table 3. Terms used for measuring the performance metrics

Term	Meaning
True Positive (TP)	Healthy image classified as healthy
True Negative (TN)	Unhealthy image classified as unhealthy
False Positive (FP)	Healthy image classified as unhealthy
False Negative (FN)	Unhealthy image classified as healthy

In MCSL classification, multiple skin lesions are classified at a single level. MCSL classification is performed using both traditional machine learning and deep learning. The dataset is divided into training and testing sets with a ratio of 70:30. When Classification is done, the performance metrics are calculated for training and testing phase separately for each category, i.e. healthy, benign, malignant, and eczema. As in MCSL algorithm, all classifications are performed at the same level, so multi-class confusion matrix is obtained. The results obtained for MCSL in the training and testing phase using traditional machine learning approach are shown in Table 4. While, the training and testing results using deep learning are given in Table 5.

Table 4. Sensitivity, specificity, precision and accuracy for healthy, benign, malignant and eczema class achieved in training and testing phase using traditional machine learning

Training Phase				
	Healthy	**Benign**	**Malignant**	**Eczema**
Sensitivity	97.07	78.24	57.71	81.26
Specificity	98.27	87	91.01	92.6
Precision	98.72	60.71	82.66	94.16
Accuracy	97.91	85.21	80.65	89.94
Testing Phase				
Sensitivity	92.37	71.69	56.18	76.5

Specificity	99.04	87.36	88.18	91.29
Precision	97.32	93.22	66.2	92.01
Accuracy	97.21	84.53	78.84	87.56

Table 5. Sensitivity, specificity, precision and accuracy for healthy, benign, malignant and eczema class achieved in training and testing phase using deep learning

Training Phase				
	Healthy	**Benign**	**Malignant**	**Eczema**
Sensitivity	100	99.5	100	100
Specificity	100	100	99.83	100
Precision	100	99.83	100	100
Accuracy	100	99.88	99.88	100
Testing Phase				
Sensitivity	99.22	81.4	91.47	95.74
Specificity	100	97.16	92.64	100
Precision	99.74	94	97.02	98.59
Accuracy	99.81	93.22	92.34	98.93

In MCML classification, the classification problem is divided into different levels. The results achieved by traditional machine learning for MCML classification for each level are given in Table 6, Table 7, and Table 8. While, the results achieved by deep learning for each level of MCML classification is shown in Table 9, Table 10, and Table 11.

Table 6. Sensitivity, specificity, precision and accuracy of healthy vs unhealthy using traditional machine learning

	Training	**Testing**
Sensitivity	100	94.92
Specificity	100	97.57
Precision	100	98.23
Accuracy	100	96.89

Table 7. Sensitivity, specificity, precision and accuracy of melanoma vs eczema using traditional machine learning

	Training	Testing
Sensitivity	98.81	94.32
Specificity	96.18	83.41
Precision	98.23	92.69
Accuracy	97.89	99.5

Table 8. Sensitivity, specificity, precision and accuracy of benign vs malignant using traditional machine learning

	Training	Testing
Sensitivity	88.34	70.19
Specificity	89.85	77.37
Precision	89.71	77.18
Accuracy	89.1	73.62

Table 9. Sensitivity, specificity, precision and accuracy for healthy vs unhealthy using deep learning

	Training	Testing
Sensitivity	100	100
Specificity	100	99.23
Precision	100	99.22
Accuracy	100	99.61

Table 10. Sensitivity, specificity precision and accuracy for melanoma vs eczema using deep learning

	Training	Testing
Sensitivity	99.83	100
Specificity	100	98.47
Precision	100	98.45

Accuracy	99.22	99.22

Table 11. Sensitivity, specificity precision and accuracy for benign vs malignant using deep learning

	Training	**Testing**
Sensitivity	99.88	92.93
Specificity	94.94	76.73
Precision	94.68	71.32
Accuracy	97.62	82.95

As mentioned earlier, validation dataset is kept aside to check the classification accuracy of the trained classifiers. The comparison of MCSL and MCML algorithm is graphically shown in Figure 17. From the figure, it is obvious that the accuracy of the traditional machine learning approach, achieved by the MCSL algorithm is 61.64% which less than the accuracy achieved by MCML algorithm that reaches 63.79%. Using deep learning approach, the MCML algorithm again performs better than MCSL. The classification accuracy achieved by MCML algorithm is 96.47%, and the classification accuracy achieved by the MCSL algorithm is 96.03%. Although using deep learning, the accuracy difference between MCML and MCSL algorithm is less, but still, this matters.

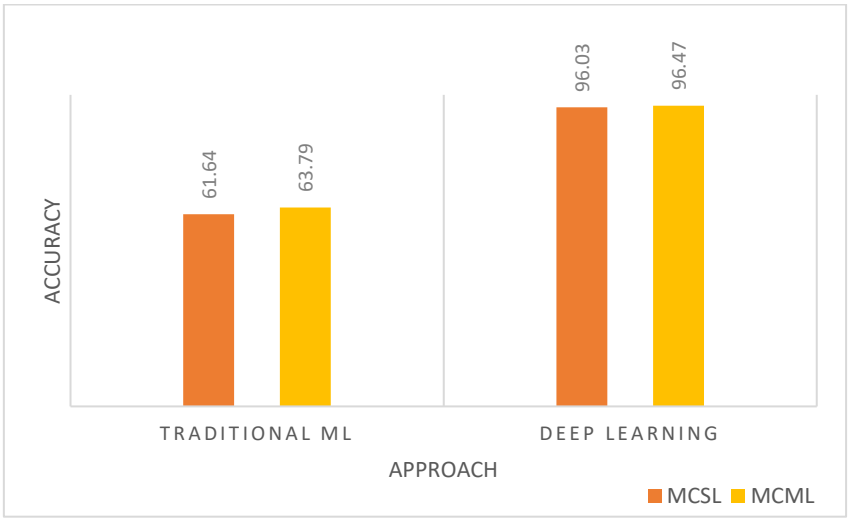

Figure 17. Comparison of MCSL and MCML using traditional machine learning and deep learning

The proposed algorithms were developed using MATLAB 2018a. Algorithms are performed on an Intel(R) Core(TM) i7-4770 CPU @ 3.40GHz with 16 GB of RAM, running Microsoft Windows 10 Enterprise 64 bits. Using traditional machine learning approach, the average time consumed in pre-processing step for each image is around 3 sec, segmentation step takes 0.95s, 0.27s are taken by

feature extraction step and classification is performed in only 6s. Total training and testing time required for traditional machine learning and deep learning approach is given in Table 12.

Table 12. Comparison of training and testing time for traditional machine learning approach and deep learning

Training Time		
	Traditional Machine Learning	**Deep Learning**
MCSL	181.13 min	35.76 min
MCML	181.25 min	53.22 min
Testing Time		
	Traditional Machine Learning	**Deep Learning**
MCSL	3.077 s	0.032 s
MCML	3.282 s	0.16 s

The training and testing time of the traditional machine learning approach is higher than the deep learning approach because of pre-processing, segmentation, feature extraction and classification step whereas in deep learning these steps are not required. Deep learning time is also high because we perform our experiments on CPU. Computational time can be reduced if experiments are performed on GPU. Although, MCML classification take more time but yields high accuracy.

While comparing the traditional machine learning and deep learning approach for multi-class skin lesion classification, deep learning always performs better except training phase of healthy vs unhealthy classification where their performance is same. The comparison of traditional machine learning and deep learning concerning MCSL and MCML algorithm is graphically presented in Figure 18, Figure 19, Figure 20, and Figure 21.

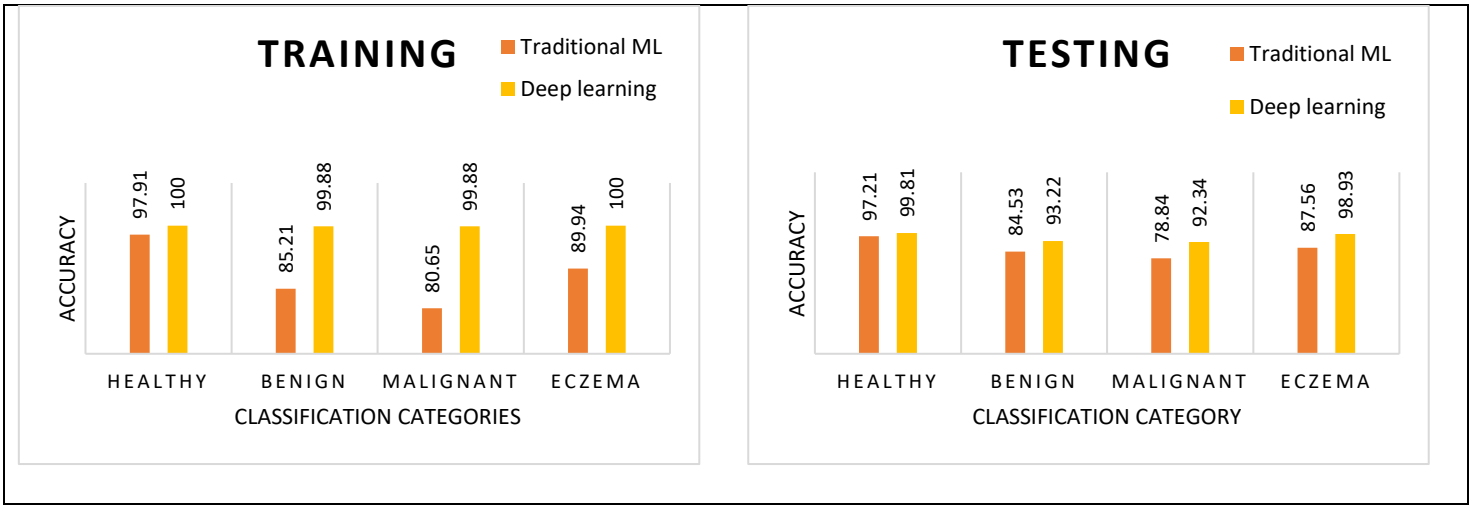

Figure 18. Comparison of traditional machine learning and deep learning for MCSL classification

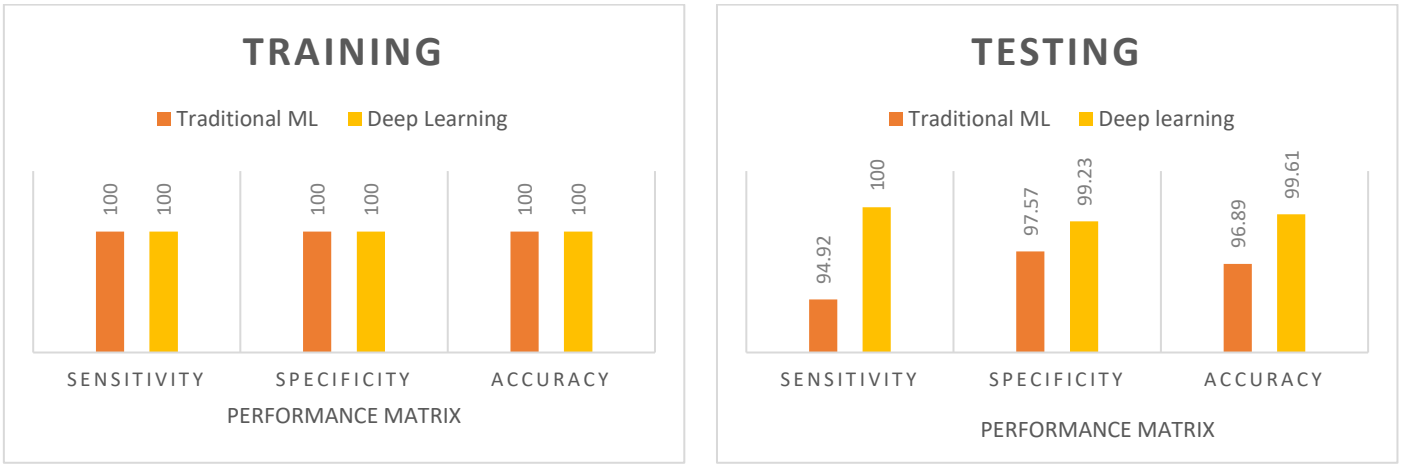

Figure 19. Comparison of traditional machine learning and deep learning for MCML (healthy vs unhealthy) classification

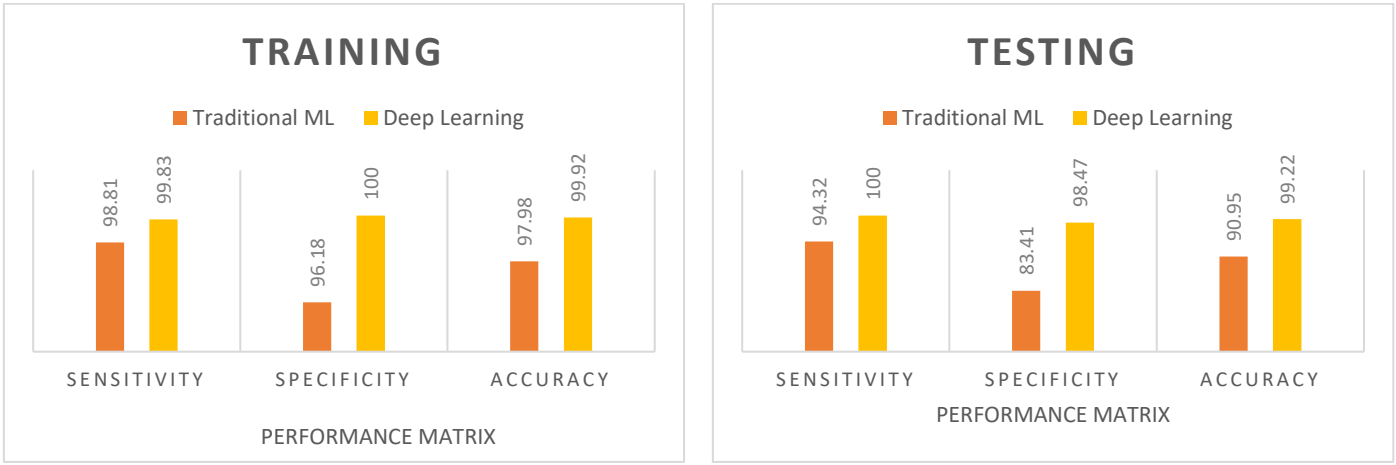

Figure 20. Comparison of traditional machine learning and deep learning for MCML (melanoma vs eczema) classification

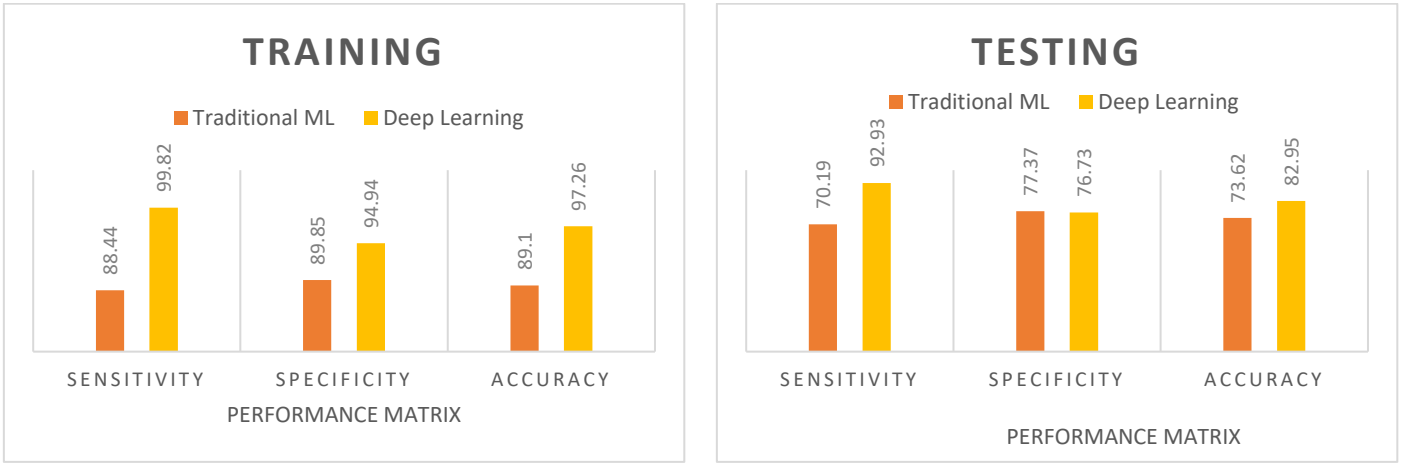

Figure 21. Comparison of traditional machine learning and deep learning for MCML (benign vs malignant) classification

This paper also compares the results of the proposed MCML algorithm with existing literature, and comparison is available in Figure 22. From the figure, it can be seen that the proposed technique outperforms most of the existing techniques for skin lesions classification. The classification accuracy achieved by Nasir et al. is better than the proposed work, but our work uses more images and classify more diseases.

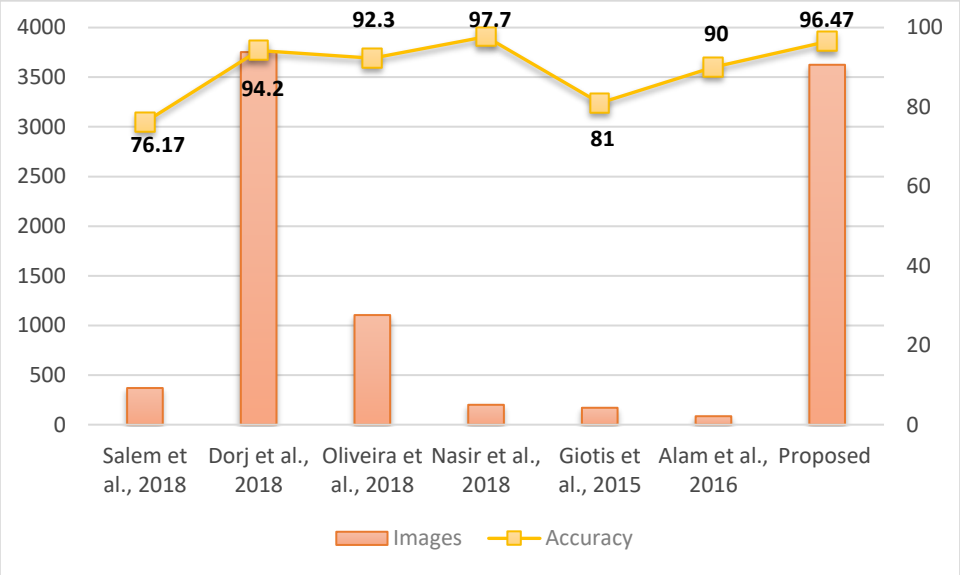

Figure 22. Comparison of proposed and existing techniques

6. Conclusion

This paper presented an investigation into the development of an intelligent multi-class multi-level (MCML) classification algorithm to classify multiple skin diseases. The developed intelligent diagnosis scheme is expected to help the users and skin specialists in early skin lesion assessment. The proposed scheme is implemented using two approaches, traditional machine learning and deep learning. Pre-processing, segmentation, feature extraction and classification steps are involved in traditional machine learning approach. Noise is removed in the pre-processing step and hybrid technique is implemented to get the region of interest in segmentation step. Colour and texture features are extracted in feature extraction step, and image is classified in the classification step. Transfer learning is used for deep learning approach and learns directly from the images. The proposed algorithm is compared with multi-class single-level classification algorithm, and high accuracy is achieved by MCML algorithm in both traditional machine learning and deep learning approach.

It is worth emphasising that previous studies have a limitation in the number of diseases they considered and the features they used for classification (Übeyli, 2008, 2009; Güvenir & Emeksiz, 2000; Abdi & Giveki, 2013; Kumar et al., 2016;) . On the contrary, the present study focuses on investigation into the development of an intelligent digital diagnosis system, where the limitation of the previous studies is overcame. The proposed algorithm is trained, tested and validated using 3672 images and 96.47% accuracy is achieved using convolutional neural network.

Future studies regarding multi-class skin lesion classification could be extended through incorporating more diseases with an objective to develop a mobile enabled expert system for the remote areas where there is no or very limited diagnosis facilities are available.

Acknowledgements.

This research work is funded by Erasmus Mundus FUSION (Featured eUrope and South asIa mObility Network) project Grant reference number:2013-32541/001001.

Conflict of interest.

The authors report no conflict of interest.

References

5 Most Common Skin Disorders. (2017). Retrieved June 1, 2018, from http://www.foxnews.com/story/2009/12/15/5-most-common-skin-disorders.html

Abbas, Q., Celebi, M. E., & Fondón, I. (2011). Biomedical Signal Processing and Control Hair removal methods : A comparative study for dermoscopy images. *Biomedical Signal Processing and Control*, *6*(4), 395–404. https://doi.org/10.1016/j.bspc.2011.01.003

Abdi, M. J., & Giveki, D. (2013). Automatic detection of erythemato-squamous diseases using PSO-SVM based on association rules. *Engineering Applications of Artificial Intelligence*, *26*(1), 603–608. https://doi.org/10.1016/j.engappai.2012.01.017

Abuzaghleh, O., Barkana, B. D., & Faezipour, M. (2014). Automated skin lesion analysis based on color and shape geometry feature set for melanoma early detection and prevention. In *2014 IEEE Long Island Systems, Applications and Technology Conference, LISAT 2014*. https://doi.org/10.1109/LISAT.2014.6845199

Afifi, M. (2017). 11K Hands: Gender recognition and biometric identification using a large dataset of hand images. *arXiv Preprint arXiv:1711.04322*. Retrieved from http://arxiv.org/abs/1711.04322

Alam, N., Tabassum, T., Munia, K., Tavakolian, K., Mackinnon, N., Fazel-rezai, R., & Member, S. (2016). Automatic Detection and Severity Measurement of Eczema Using Image Processing. In *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1365–1368).

British Skin Foundation. (2018). Retrieved May 20, 2018, from http://www.britishskinfoundation.org.uk/Home.aspx

Burdick, J., Marques, O., Weinthal, J., & Furht, B. (2017). Rethinking Skin Lesion Segmentation in a Convolutional Classifier. *Journal of Digital Imaging*, 1–6. https://doi.org/10.1007/s10278-017-0026-y

Çataloluk, H., & Kesler, M. (2012). A Diagnostic Software Tool for Skin Diseases with Basic and Weighted K-NN. In *2012 International Symposium on Innovations in Intelligent Systems and Applications*.

Celebi, M. E., Aslandogan, Y. a., & Bergstresser, P. R. (2005). Unsupervised border detection of skin lesion images. *International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume II*, *2*, 1–6. https://doi.org/10.1109/ITCC.2005.283

Chang, C.-Y., & Liao, H.-Y. (2011). Automatic Facial Skin Defects Detection and Recognition System. *2011 Fifth International Conference on Genetic and Evolutionary Computing*, 260–263. https://doi.org/10.1109/ICGEC.2011.67

Chang, C. L., & Chen, C. H. (2009). Applying decision tree and neural network to increase quality of dermatologic diagnosis. *Expert Systems with Applications*, *36*(2 part 2), 4035–4041. https://doi.org/10.1016/j.eswa.2008.03.007

Codella, N. C. F., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., … Halpern, A. (2016). Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). In *arXiv:1605.01397* (pp. 3–7).

De Guzman, L. C., Maglaque, R. P. C., Torres, V. M. B., Zapido, S. P. A., & Cordel, M. O. (2015). Design and Evaluation of a Multi-model, Multi-level Artificial Neural Network for Eczema Skin Lesion Detection. *2015 3rd International Conference on Artificial Intelligence, Modelling and Simulation (AIMS)*, 42–47. https://doi.org/10.1109/AIMS.2015.17

Dermatologist, B. A. of. (2015). The global challenge for skin health. *Bristish Journal of Dermatloogy*, *172*, 1469–1472. https://doi.org/10.1111/bjd.13854

DermIS. (2018). DermIS. Retrieved June 29, 2017, from http://www.dermis.net/dermisroot/en/home/index.htm

DermNZ. (2018). Retrieved January 13, 2018, from https://www.dermnetnz.org/image-library/

DermQuest Image Library. (2018). DermQuest Image Library. Retrieved January 12, 2018, from https://www.dermquest.com/image-library/

Dorj, U., Lee, K., Choi, J., & Lee, M. (2018). The skin cancer classification using deep convolutional neural network. *Multimedia Tools and Applications*, *77*(8), 9909–9924. https://doi.org/https://doi.org/10.1007/s11042-018-5714-1

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118. https://doi.org/10.1038/nature21056

Freedberg, K. A., Geller, A. C., Miller, D. R., Lew, R. A., & Koh, H. K. (1999). Screening for malignant melanoma: A cost-effectiveness analysis. *Journal of the American Academy of Dermatology*, *41*(5 I), 738–745. https://doi.org/10.1016/S0190-9622(99)70010-1

Ganster, H., Pinz, A., Rohrer, R., Wildling, E., Binder, M., & Kittler, H. (2001). Automated melanoma recognition. *IEEE Transactions on Medical Imaging*, *20*(3), 233–239. https://doi.org/10.1109/42.918473

George, Y., Aldeen, M., & Garnavi, R. (2015). Skin Hair Removal for 2D Psoriasis Images. In *International Conference on Digital Image Computing: Techniques and Applications*.

Giotis, I., Molders, N., Land, S., Biehl, M., Jonkman, M. F., & Petkov, N. (2015). MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert Systems with Applications*, *42*(19), 6578–6585. https://doi.org/10.1016/j.eswa.2015.04.034

Güvenir, H. A., & Emeksiz, N. (2000). Expert system for the differential diagnosis of erythemato-squamous diseases. *Expert Systems with Applications*, *18*(1), 43–49. https://doi.org/10.1016/S0957-4174(99)00049-4

Hameed, N., Hassan, K. A., & Hossain, M. A. (2016). A comprehensive survey on image-based computer aided diagnosis systems for skin cancer. In *2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA)* (pp. 205–214). IEEE. https://doi.org/10.1109/SKIMA.2016.7916221

Hay RJ, Augustin M, Griffiths CE, S. W. (2015). The global challenge for skin health. *British Journal of Dermatology*, *172*(6), 1469–1472. https://doi.org/10.1111/bjd.13854

Ilter, N., & Guvenir, H. A. (1998). UCI Machine Learning Repository: Dermatology Data Set. Retrieved April 27, 2017, from https://archive.ics.uci.edu/ml/datasets/Dermatology

Jain, S., Jagtap, V., & Pise, N. (2015). Computer aided melanoma skin cancer detection using image processing. *Procedia Computer Science*, *48*(C), 736–741. https://doi.org/10.1016/j.procs.2015.04.209

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, *6*(5), 429–449.

Jaworek-korjakowska, J., & Kleczek, P. (2018). eSkin : Study on the Smartphone Application for Early Detection of Malignant Melanoma. *Wireless Communications and Mobile Computing*, *2018*. https://doi.org/https://doi.org/10.1155/2018/5767360

Karimkhani, C., Dellavalle, R. P., & Coffeng, L. E. (2017). Global Skin Disease Morbidity and Mortality An Update From the Global Burden of Disease Study 2013. *Jama Dermatology*, *153*(5), 406–412. https://doi.org/10.1001/jamadermatol.2016.5538

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, 1–9. https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007

Kumar, V. B., Kumar, S. S., & Saboo, V. (2016). Dermatological Disease Detection using Image Processing and Artificial Neural Network. *International Conference on Artificial Intelligence and Pattern Recognition (AIPR)*, 88–93. https://doi.org/10.1109/ICAIPR.2016.7585217

Lee, T., Ng, V., Gallagher, R., Coldman, A., & McLean, D. (1997). Dullrazor: A software approach to hair removal from images. *Computers in Biology and Medicine*, *27*(6), 533–543. https://doi.org/10.1016/S0010-4825(97)00020-6

Maglogiannis, I., & Doukas, C. N. (2009). Overview of advanced computer vision systems for skin lesions characterization. *IEEE Transactions on Information Technology in Biomedicine*, *13*(5), 721–723. https://doi.org/10.1109/TITB.2009.2017529

Maglogiannis, I., Ieee, S. M., & Delibasis, K. (2015). Hair Removal on Dermoscopy Images. In *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 2960–2963).

Mathworks. (2018). Nearest Neighbor, Bilinear, and Bicubic Interpolation Methods. Retrieved May 30, 2018, from https://uk.mathworks.com/help/vision/ug/interpolation-methods.html

MATLAB. (2017). (GLCM) - MATLAB & Simulink. Retrieved April 14, 2018, from https://uk.mathworks.com/help/images/ref/graycomatrix.html

Melanoma - Skin Dermatologists. (2016). Retrieved May 10, 2017, from http://www.skindermatologists.com/melanoma.html

Mendonca, T., Ferreira, P. M., Marques, J. S., Marcal, A. R. S., & Rozeira, J. (2013). PH2 - A dermoscopic image database for research and benchmarking. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* (pp. 5437–5440). https://doi.org/10.1109/EMBC.2013.6610779

Nanni, L. (2006). An ensemble of classifiers for the diagnosis of erythemato-squamous diseases. *Neurocomputing*, *69*(7–9), 842–845. https://doi.org/10.1016/j.neucom.2005.09.007

Nasir, M., Attique Khan, M., Sharif, M., Lali, I. U., Saba, T., & Iqbal, T. (2018). An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach. *Microscopy Research and Technique*, *81*(6), 528–543. https://doi.org/10.1002/jemt.23009

Nugroho, H., Ahmad Fadzil, M. H., Shamsudin, N., & Hussein, S. H. (2013). Computerised image analysis of vitiligo lesion: Evaluation using manually defined lesion areas. *Skin Research and Technology*, *19*(1), 72–77. https://doi.org/10.1111/j.1600-0846.2011.00610.x

Oliveira, R. B., Marranghello, N., Pereira, A. S., & Tavares, J. M. R. S. (2016). A computational approach for detecting pigmented skin lesions in macroscopic images. *Expert Systems with Applications*, *61*, 53–63. https://doi.org/10.1016/j.eswa.2016.05.017

Oliveira, R. B., Pereira, A. S., Manuel, J., & Tavares, R. S. (2016). Computational methods for the image segmentation of pigmented skin lesions : A review. *Computer Methods and Programs in Biomedicine*, *131*, 127–141. https://doi.org/10.1016/j.cmpb.2016.03.032

Oliveira, R. B., Pereira, A. S., & Tavares, J. M. R. S. (2018). Computational diagnosis of skin lesions from dermoscopic images using combined features. *Neural Computing and Applications*, 1–21. https://doi.org/10.1007/s00521-018-3439-8

Picardi, A. (2013). Suicide risk in skin disorders. *Clinics in Dermatology*, *31*, 47–56. https://doi.org/10.1016/j.clindermatol.2011.11.006

Rubegni, P., Cevenini, G., Burroni, M., Perotti, R., Dell'Eva, G., Sbano, P., … Andreassi, L. (2002). Automated diagnosis of pigmented skin lesions. *International Journal of Cancer*, *101*(6), 576–580. https://doi.org/10.1002/ijc.10620

Salem, C., Azar, D., & Tokajian, S. (2018). An Image Processing and Genetic Algorithm-based Approach for the Detection of Melanoma in Patients. *Methods of Information in Medicine*, *57*(1), 74–80. https://doi.org/10.3412/ME17-01-0061

Shamsul Arifin, M., Golam Kibria, M., Firoze, A., Ashraful Amini, M., & Yan, H. (2012). Dermatological disease diagnosis using color-skin images. In *International Conference on Machine Learning and Cybernetics* (Vol. 5, pp. 1675–1680). https://doi.org/10.1109/ICMLC.2012.6359626

Siegel, R. L., Miller, K. D., & Jemal, A. (2018). Cancer Statistics , 2018, *66*(1), 7–30. https://doi.org/10.3322/caac.21332.

Sultana, A., Dumitrache, I., Vocurek, M., & Ciuc, M. (2014). Removal of artifacts from dermatoscopic images. In *IEEE International Conference on Communications*. https://doi.org/10.1109/ICComm.2014.6866757

Toossi, M. T. B., Pourreza, H. R., Zare, H., Sigari, M. H., Layegh, P., & Azimi, A. (2013). An effective hair removal algorithm for dermoscopy images. *Skin Research and Technology*, *19*(3), 230–235. https://doi.org/10.1111/srt.12015

Übeyli, E. D. (2008). Multiclass support vector machines for diagnosis of erythemato-squamous diseases. *Expert Systems with Applications*, *35*(4), 1733–1740. https://doi.org/10.1016/j.eswa.2007.08.067

Übeyli, E. D. (2009). Combined neural networks for diagnosis of erythemato-squamous diseases. *Expert Systems with Applications*, *36*(3 PART 1), 5107–5112. https://doi.org/10.1016/j.eswa.2008.06.002

Vasconcelos, C. N., & Vasconcelos, B. N. (2017). Experiments using deep learning for dermoscopy image analysis. *Pattern Recognition Letters*, *0*, 1–9. https://doi.org/10.1016/j.patrec.2017.11.005

Victor, A., & Ghalib, M. (2017). Automatic Detection and Classification of Skin Cancer. *International Journal of Intelligent Engineering and Systems*, *10*(3), 444–451. https://doi.org/10.22266/ijies2017.0630.50

World Health Organization. (2018). WHO | Skin cancers. Retrieved May 13, 2018, from http://www.who.int/uv/faq/skincancer/en/index1.html

Xie, J., & Wang, C. (2011). Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases. *Expert Systems with Applications*, *38*(5), 5809–5815. https://doi.org/10.1016/j.eswa.2010.10.050

Zakeri, A., & Hokmabadi, A. (2018). Improvement in the diagnosis of melanoma and dysplastic lesions by introducing ABCD-PDT features and a hybrid classifier. *Biocybernetics and Biomedical Engineering*, 1–11. https://doi.org/10.1016/j.bbe.2018.03.005

Zheng, A. (2015). Evaluating Machine Learning Models. Retrieved May 19, 2018, from https://www.oreilly.com/ideas/evaluating-machine-learning-models/page/4/offline-evaluation-mechanisms-hold-out-validation-cross-validation-and-bootstrapping